Evaluation of Statistically Based Cloudiness Parameterizations Used in Climate Models

KUAN-MAN XU AND DAVID A. RANDALL

Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

(Manuscript received 29 December 1995, in final form 15 April 1996)

ABSTRACT

Existing cloudiness parameterizations based on specified probability distribution functions (PDFs) and large-scale relative humidity (RH) in climate models are evaluated with data produced from explicit simulations of observed tropical cloud systems and subtropical stratocumuli. PDF-based parameterizations were originally intended for use in cloud-resolving models, where fractional cloudiness is only associated with turbulence-scale motion. It is demonstrated with simulated data that most PDF-based parameterizations are not adequate for predicting fractional cloudiness in climate models because their performance is dependent upon the cloud regimes. Modifications to some PDF-based formulations are suggested, especially with regard to the inclusion of skewness of conservative variables. The skewness factors are found to be highly dependent upon which scales of motion coexist within a grid cell. RH-based parameterizations are not readily supported due to a wide range of variations of clear-region averaged RHs with height and the grid size of climate models, as well as their wide range of variations at a given height.

1. Introduction

Within the framework of the prognostic approach for cloud formation and dissipation processes as pioneered by Sundqvist (1978), there have been many existing cloudiness parameterizations used in global climate models (GCMs; e.g., Sundqvist 1978; Smith 1990; Le Treut and Li 1988; Ricard and Royer 1993). Most of them use simple assumed or predicted probability distribution functions (PDFs) for certain variables, such as the total water mixing ratio q_w and liquid-water potential temperature θ_l, to estimate the fractional cloudiness. These can be called *statistically based* cloudiness parameterizations. Some of these parameterizations can be formulated as simple relations between cloudiness and large-scale variables such as relative humidity (Sundqvist 1978; Le Treut and Li 1988; Smith 1990; Sundqvist et al. 1989), similar to those used in the diagnostic approach (Xu and Krueger 1991). Such PDF-based parameterizations were originated by Sommeria and Deardorff (1977) and Mellor (1977), who developed a subgrid cloudiness parameterization intended for use in high-resolution, cloud-resolving models. The basic idea behind the Sommeria–Deardorff–Mellor parameterization is that a pair of moist conservative variables, such as θ_l and q_w, undergoing subgrid-scale fluctuations, have a joint Gaussian PDF, as sketched in Fig. 1. The left side of the dashed line in the sketch represents subsaturated air, and the right side represents saturated air. By integrating the joint PDF over the saturated region, the cloud amount can be determined.

A requirement for the PDF-based approach is knowledge of the subgrid distribution of the moist conservative variables. In other words, the main limitation of such PDF-based parameterizations is that the distribution is generally assumed (Table 1). For example, Sundqvist's (1978) parameterization assumed 100% RH in the cloudy region and a constant threshold value in the clear region. The fractional cloudiness is thus dependent only on the cell-averaged RH and the threshold RH for zero cloudiness, regardless of the amount of cloud water/ice in the cell. Sundqvist et al. (1989) assumed that the threshold RH also depends on the fractional cloudiness. Le Treut and Li (1988, 1991) used a uniform distribution of q_w to estimate the fractional cloudiness. They were not sure how to determine the magnitude of the PDF. On the other hand, Ricard and Royer (1993) closely followed the Sommeria–Deardorff–Mellor approach, in which the variances of moist conservative variables were determined by a turbulence parameterization (Yamada and Mellor 1979) and a joint Gaussian PDF of θ_l and q_w was assumed. The adequacy of such a turbulence parameterization for macroscopic convective turbulence within stratiform clouds was not, however, addressed by Ricard and Royer (1993). Smith (1990) assumed a symmetric triangular joint PDF for θ_l and q_w. With additional assumptions, he deduced a formula for fractional cloudiness as a function of the cell-averaged RH and an assumed threshold RH.

Corresponding author address: Dr. Kuan-Man Xu, Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523.
E-mail: kmxu@cumulus.atmos.colostate.edu

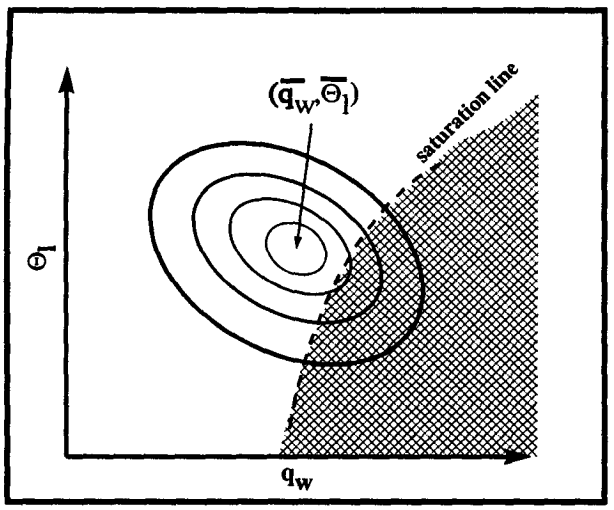

FIG. 1. A sketch illustrating the Sommeria–Deardorff–Mellor approach to cloud amount parameterization. The hatched area is saturated. The ovals represent contours of the joint probability function of liquid-water potential temperature θ_l and total water mixing ratio q_w.

Tiedtke (1993) introduced a fully prognostic cloud parameterization, which consists of prognostic equations for the cloud amount and the cloud mass. The evolution of cloud amount and water/ice content is thus fully determined by advective processes and the sources and sinks due to diabatic processes. The sources and sinks of cloud amount were parameterized for major cloud types, such as frontal, cumulus, anvil/cirrus and boundary layer clouds, in plausible, albeit speculative, ways. The explicit link between anvil/cirrus clouds with penetrative cumulus convection was also included.

Sasamori (1975) hypothesized that clouds in a certain layer were generated by an ensemble of air parcels vertically displaced at random. A Gaussian distribution for the displaced lengths of air parcels was assumed. The cloud amount was determined as the ensemble average of those parcels whose lifting condensation level lies below the top edge of a layer. Condensation and precipitation can also be diagnosed under the statistical equilibrium between subgrid- and grid-scale motions.

One of the drawbacks of RH-based formulations (Sundqvist 1978; Sundqvist et al. 1989; Smith 1990) is that the cell-averaged RH can be low in the presence of active deep convection because of the drying and warming effects of compensating subsidence (Yanai et al. 1973; Arakawa and Schubert 1974), while the coverage of the associated anvils may not be small. Moreover, in the absence of deep convection the cell-averaged RH might not be a good measure of stratiform cloudiness if there is no unique choice of threshold RHs.

A fundamental question is why should the PDF-based formulations be expected to work on the GCM grid size. PDF-based parameterizations were derived from turbulence principles and were originally intended for use in high resolution cloud models with a grid size of less than 1 km^2. The unresolved motions in cloud models are basically three-dimensional turbulence. In GCMs, however, the unresolved motions range from turbulence through cloud scale to mesoscale. The PDFs of a given variable or a pair of variables are undoubtedly different for these two types of models. Therefore, a basic understanding of the PDFs of certain variables in a GCM grid cell is needed to verify the aforementioned formulations. Bougeault (1981) performed an evaluation of three types of PDFs with data from a large-eddy simulation but little attention has been paid subsequently to evaluating this approach for different cloud regimes.

A main objective of the present study is to evaluate the PDF-based approach in existing cloudiness parameterizations and to suggest modifications to some existing formulations. Another objective is to further evaluate the RH-based parameterizations by specifically examining the existence of unique RH thresholds.

2. The approach

The method adopted in the present study follows Xu and Krueger (1991) who used cloud ensemble simulations to investigate the utility of several empirical parameterizations based on different "predictors" such

TABLE 1. Characteristics of some cloudiness parameterizations.

Parameterization	Predictor(s)	PDF	Note
Sasamori (1975)	vertical path length of air parcels	Gaussian	also diagnosing the cloud water content
Sundqvist (1978)	RH	RH = 1 in cloudy region and RH_0 in clear region	RH_0 depends also on cloudiness in Sundqvist et al.
Le Treut and Li (1988)	q_w	uniform	uncertain magnitude of PDF
Smith (1990)	θ_l and q_w	joint symmetric triangular	assumptions lead to a RH-type formula
Tiedtke (1993)	prognostic equation with sources/sinks	none	sources/sinks differ among major cloud types
Ricard and Royer (1993)	θ_l and q_w	joint Gaussian	requires a parameterization for macroscopic convective turbulence

as the large-scale RH, the large-scale vertical motion, the surface precipitation rate, and the convective mass flux. All of these predictors showed some utility; for example, the large-scale RH was useful for predicting stratiform cloudiness and the convective mass flux was helpful for determining convective cloudiness. None of the predictors, however, were sufficiently skillful for quantitative applications in GCMs.

Due to the lack of observations of the detailed, statistical properties of cloud fields, a cloud ensemble model (CEM) is used to generate datasets with pertinent cloud variables to evaluate statistically based cloudiness parameterizations in this study. CEMs can resolve the structures of individual clouds over a spatial domain large enough to contain many clouds and can run long enough to include many cloud life cycles (e.g., Soong and Tao 1980; Lipps and Hemler 1986; Krueger 1988; Gregory and Miller 1989). The two-dimensional UCLA CEM (Krueger 1988) is used for the simulations described below.

Simulations of tropical cloud systems observed during the Global Atmospheric Research Program's (GARP) Atlantic Tropical Experiment (GATE) and subtropical stratus and stratocumulus during the Atlantic Stratocumulus Transition EXperiment (ASTEX) are analyzed for this study. The GATE simulation was described by Xu and Randall (1996a), while the ASTEX simulation was briefly presented by Xu and Randall (1996b). In both simulations, the observed large-scale vertical velocity, horizontal advections of heat and moisture, and sea surface temperature are imposed to the model uniformly in the horizontal. The GATE simulation was run from 1 to 18 September 1974, almost the entire length of GATE Phase III. The horizontal domain size is 512 km with a grid size of 2 km. The domain depth extends to approximately 19 km with a stretched spacing from 100 to 1000 m. The ASTEX simulation was run from 17 to 22 June 1992. The horizontal extent of the domain is 128 km with a grid size of 2 km. The vertical extent of the domain is 3.6 km with a uniform spacing of 80 m.

3. An evaluation of PDF-based formulations

a. Description of PDF-based formulations

First, a brief description of the general methodology of PDF-based formulations is given. Following Bougeault (1981), the joint PDF between liquid-water potential temperature θ_l and total water mixing ratio q_w can be reformulated in terms of a single variable s, where

$$s = a(q_w' - \alpha_1\theta_l')/2,$$

the primes denote the deviation from the mean state, and a and α_1 depend only on the mean thermodynamic variables. That is, $\alpha \equiv \partial q^*(\bar{\theta}_l)/\partial\bar{\theta}_l$ and $a \equiv [1 + L/c_p\partial q^*(\bar{\theta}_l)/\partial\bar{T}_l]^{-1}$, where the overbars denote the mean state, q^* is the saturation water-vapor mixing ratio, and T_l is the liquid-water temperature. The physical meaning of s is the distance from the actual thermodynamic state to the saturation curve (Fig. 1).

The variance of s, σ_s, is a combination of the variances of θ_l and q_w; that is,

$$\sigma_s = (a/2)[\overline{q_w'^2} + \alpha_1^2\overline{\theta_l'^2} - 2\alpha_1\overline{\theta_l' q_w'}]^{1/2}. \quad (1)$$

Let t be the normalized variable s/σ_s, $G(t)$ its probability density, and Ω_1 a dimensionless measure of the departure of the mean state from saturation, which can be expressed as

$$\Omega_1 = a[\overline{q_w} - q^*(\bar{\theta}_l)]/2\sigma_s.$$

Then the cloud fraction C_s and the liquid-water mixing ratio $\overline{q_l}$ are determined by

$$C_s = \int_{-\Omega_1}^{+\infty} G(t)dt, \quad (2)$$

$$\overline{q_l}/2\sigma_s = \int_{-\Omega_1}^{+\infty} (\Omega_1 + t)G(t)dt. \quad (3)$$

Thus for a given PDF, C_s and $\overline{q_l}$ depend only upon the normalized departure of the mean state from the saturation, Ω_1. It should be noted that the saturation state here is defined with respect to T_l. Thus, it is possible to have a supersaturated mean state.

As in Bougeault (1981), three PDFs of the joint θ_l and q_w are evaluated with the CEM data, namely,

(i) the Gaussian model

$$G(t) = \frac{1}{\sqrt{2\pi}}\exp(-t^2/2), \quad (4)$$

(ii) an exponential model

$$G(t) = \frac{1}{\sqrt{2}}\exp(-\sqrt{2}|t|), \quad (5)$$

and (iii) a simple positively skewed model

$$G(t) = H(t + 1)\exp[-(t + 1)], \quad (6)$$

where $H(t)$ is the Heaviside function. Simple algebraic expressions for C_s and $\overline{q_l}$ with these three models can be found in Bougeault (1981). The positively skewed model allows for a constant skewness factor of 2.

b. GATE data

Figure 2 shows the prediction of the domain-averaged (512 km) stratiform cloud amounts obtained with the three models, based on input data from the GATE simulation, which are time averaged over 3 h and spatially averaged over the entire domain. In this simulation, both convective and stratiform clouds are produced (Xu and Randall 1996a). A partitioning method (Xu 1995) is used to distinguish stratiform regions from convective regions. Xu and Randall (1996b)

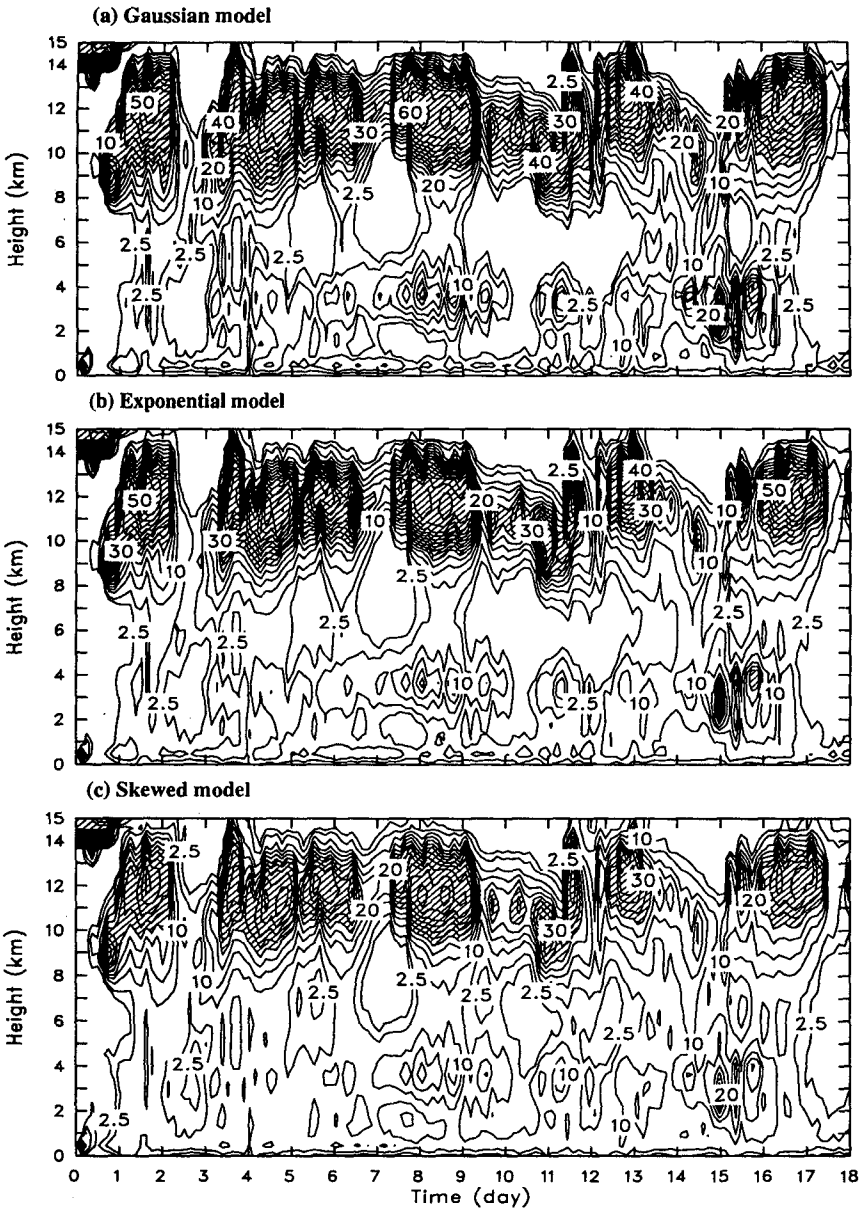

FIG. 2. Time–height cross sections of stratiform cloud amounts for a GATE simulation predicted by (a) a Gaussian model, (b) an exponential model, and (c) a skewed model. The contour interval is 5%, with an additional contour of 2.5%.

showed some results related to the partitioning. Convective clouds are excluded in the results presented for the GATE data.

All three models are successful in predicting the locations of maxima and minima [cf. with Fig. 4b in Xu and Randall (1996b)]. This is more evident from the vertical profiles of the correlation coefficient between predicted and simulated cloud amounts (Fig. 3a). However, almost all maxima are overpredicted by the three models, especially in the upper troposphere (see also Fig. 3b). The predicted maxima in the upper troposphere are located approximately 1 km higher than the simulated ones [Fig. 4b in Xu and Randall (1996b) and Fig. 3b]. The standard deviations of the predicted C_s are also much larger in the upper troposphere than those of the simulated C_s (Fig. 3c). The smallest overprediction is obtained with the skewed model (Figs. 2c, 3b, and 3c), which suggests that the PDF estimated from the simulated data is skewed (see section 4). On the other hand, small underprediction occurs in the middle troposphere with the Gaussian and exponential models, which can be related to the dry mean states in

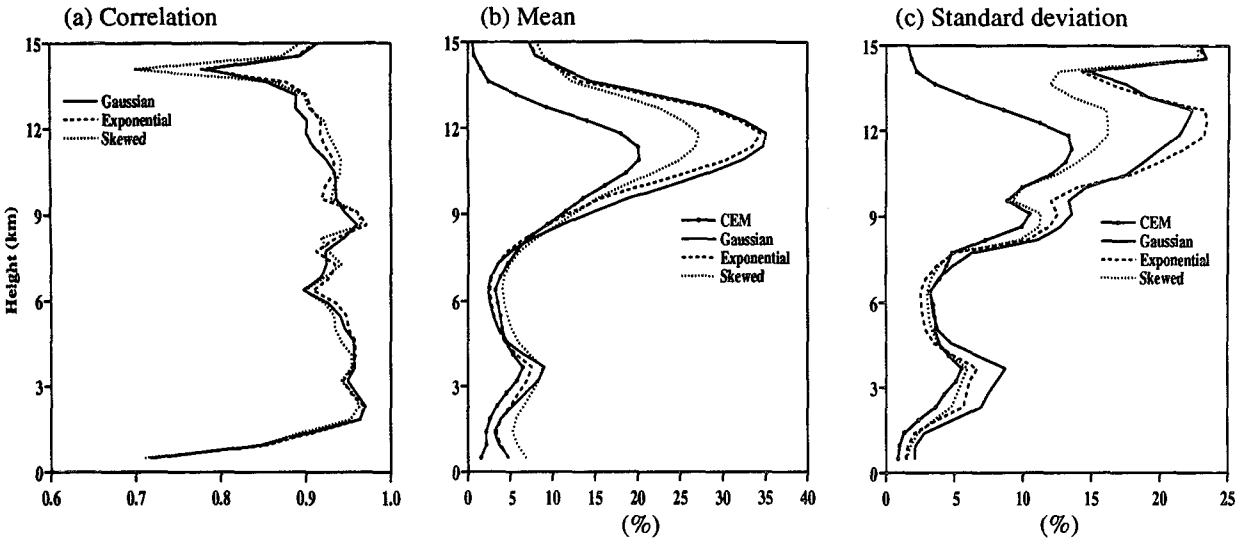

FIG. 3. Vertical profiles of (a) the correlation coefficient between simulated and predicted cloud amounts, (b) the mean, and (c) the standard deviation of cloud amounts for the 512-km domain average of a GATE simulation.

this layer. Bougeault (1982) showed that the Gaussian model underestimates the cloud amount at dry mean states and overestimates the cloud amount at moist mean states if the skewness factors of the distribution are positive, which are defined as $S_s \equiv \overline{s^3}/(\overline{s^2})^{3/2}$.

Figure 4 shows similar profiles as in Fig. 3 except for $\bar{q}_l$. Unlike C_s, the correlation coefficients for $\bar{q}_l$ are extremely high (>0.97) in the upper troposphere for all three models, while they are much lower in the middle troposphere (Fig. 4a). The mean profiles of $\bar{q}_l$ show large overestimates by the exponential and skewed models at all levels (Fig. 4b). The Gaussian model shows the best agreement in the mean $\bar{q}_l$ profile except for a higher level of maximum in the upper troposphere and small underestimates between 1 and 10 km. The standard deviations of the predicted $\bar{q}_l$ by the exponential and skewed models are also much larger than the simulated ones (Fig. 4c). Therefore, the Gaussian model appears superior to the other two models as far as $\bar{q}_l$ is concerned.

Figure 5 shows scatterplots of simulated versus predicted C_s at three selected levels for the 64-km subdomains of the GATE simulation. The horizontal axis is the simulated C_s and the vertical axis is the predicted C_s from

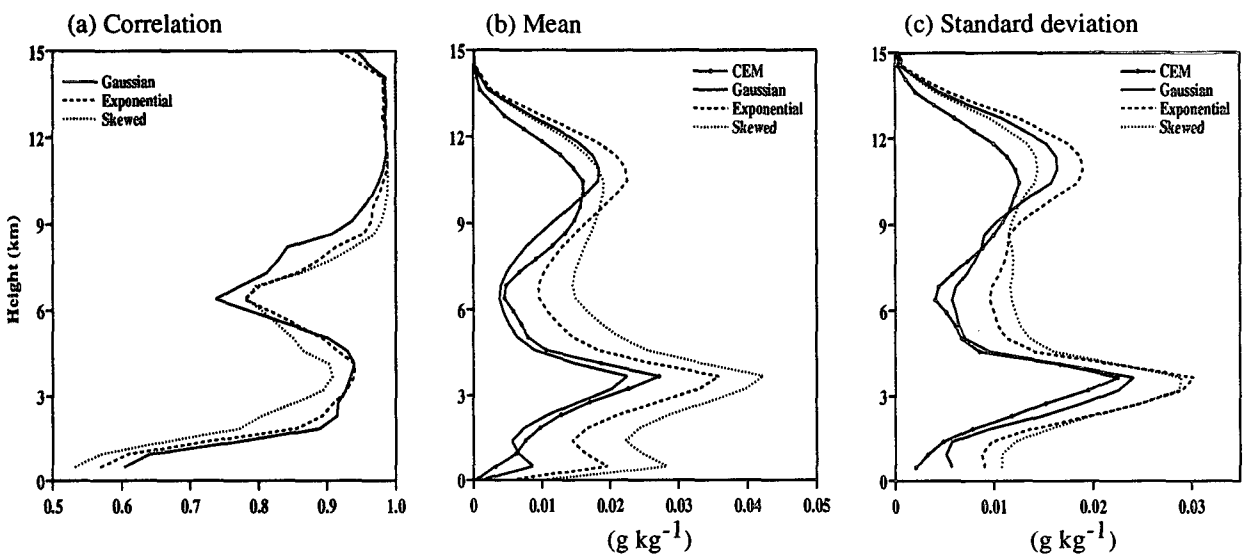

FIG. 4. Same as Fig. 3 except for cloud water + ice mixing ratio.

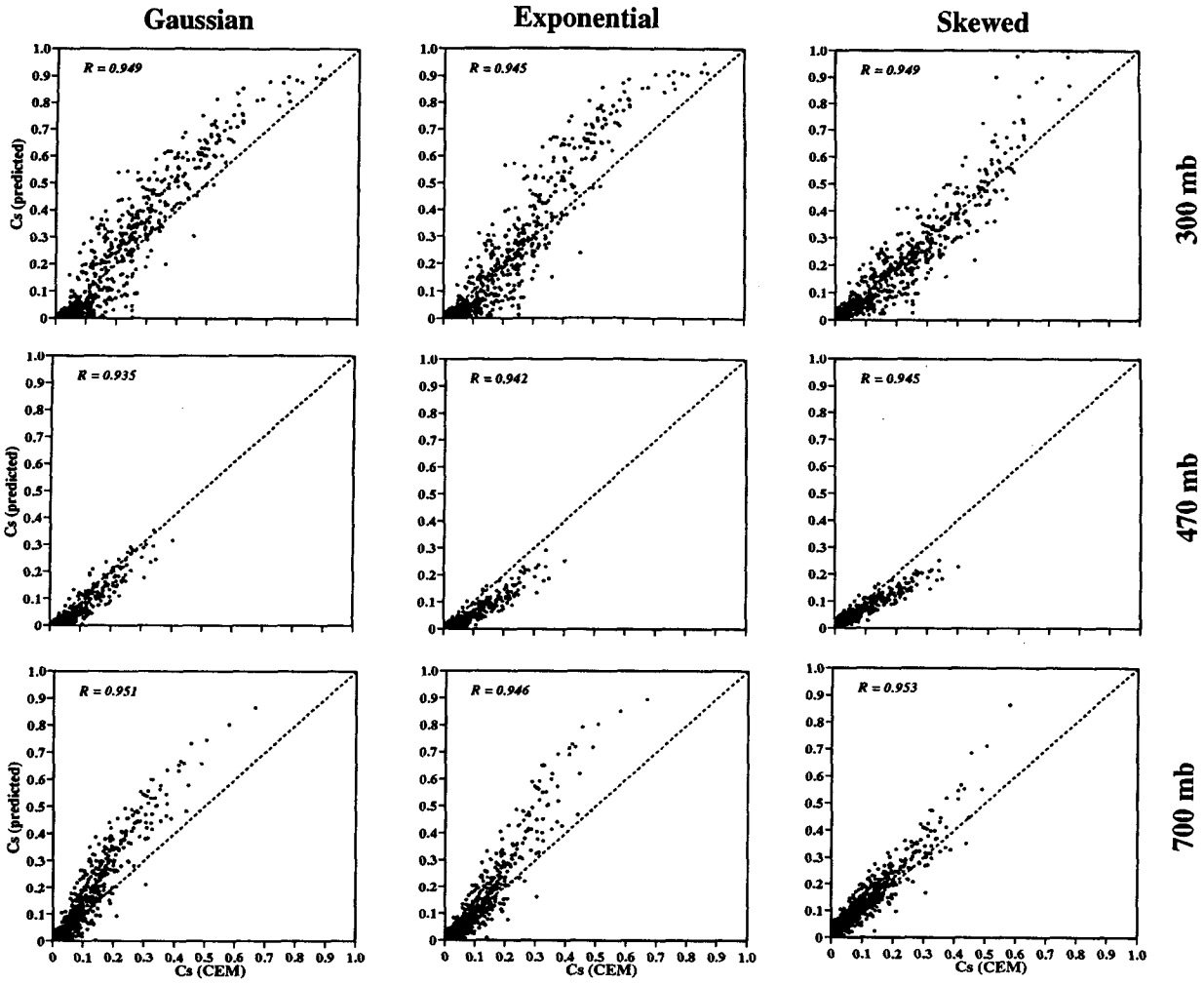

FIG. 5. Scatterplot of stratiform cloud amounts from CEM vs those predicted with the Gaussian model (left panels), the exponential model (middle panels), and the skewed model (right panels) at 300-, 470-, and 700-mb levels for the 64-km subdomain of a GATE simulation. Here R is the correlation coefficient between simulated and predicted C_s.

one of the three models. The three selected levels are representative of the low, middle, and upper troposphere (Figs. 3b and 3c). The correlation coefficients between simulated and predicted C_s are very high at the three levels, as in the 512-km average (Fig. 3a). The scatter from the "perfect" prediction (i.e., the dashed line) is, however, not negligible at all levels. The cloud amount is mostly overpredicted in the lower and upper troposphere, while it is mostly underpredicted in the middle troposphere. The overprediction is generally larger for both Gaussian and exponential models when the simulated C_s is larger, corresponding to cases with positive Ω_1 and positively skewed PDFs. This is consistent with the 512-km average results (Fig. 3). The magnitude of underprediction does not necessarily depend on the simulated C_s but occurs mostly in the low range of the C_s variation; that is, C_s is less than 0.5 and Ω_1 is negative. Thus, the skewed model is slightly superior for predicting C_s, as found by Bougeault (1981) for stratocumulus, since the skewed model gives a better prediction in the low and upper troposphere.

The features appearing in the scatterplots of simulated versus predicted $\bar{q}_l$ for the 64-km subdomains (Fig. 6) are similar to those in C_s (Fig. 5), for example, the extremely high correlation between simulated and predicted $\bar{q}_l$ at each level. There are three noticeable differences: (i) the scatter from the perfect prediction is considerably less significant in the middle and upper troposphere, (ii) the overprediction in $\bar{q}_l$ is not as systematic as that in C_s, and (iii) $\bar{q}_l$ in the middle troposphere is mostly overpredicted by the exponential and skewed models. These features are consistent with the 512-km average results (Fig. 4). In summary, C_s is more difficult to predict using the PDF-based models than is $\bar{q}_l$. These results are consistent with Bougeault (1982), who showed that the impact of skewness on $\bar{q}_l$

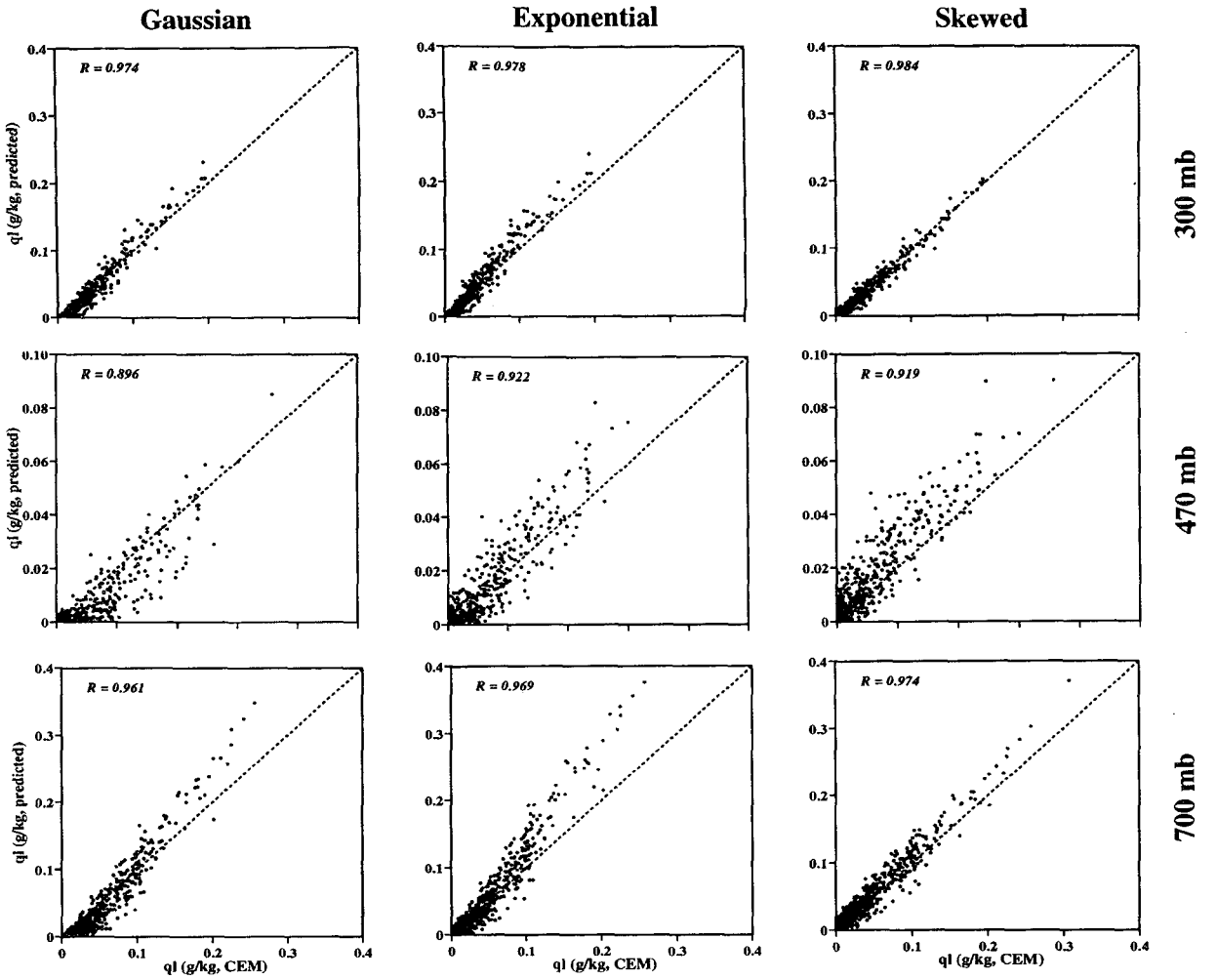

FIG. 6. Same as Fig. 5 except for simulated vs predicted liquid water mixing ratios.

is relatively less than on C_s at any mean state except for the extremely dry one.

Figure 7 shows the dependence of Gaussian PDF prediction of cloud amounts on input data, which are averaged over 512 km and 64 km, respectively. The predicted C_s from 64-km average inputs are averaged to obtain C_s for the entire domain. Figure 7 clearly shows that the performance of Gaussian PDF model improves as the horizontal-averaging distance decreases. The overprediction of C_s does, however, not diminish for the 64-km average results due to skewness of the simulated data.

c. *ASTEX data*

Figure 8 shows the differences between predicted and simulated cloud amount obtained with the three PDF models, based on domain-averaged data from the ASTEX simulation (Xu and Randall 1996b). The ASTEX data are used to show whether or not the performance of PDF models is dependent upon the cloud regimes. No partitioning between stratus and stratocumulus was attempted for the ASTEX data. With the Gaussian model, significant overprediction of C_s ($\sim$10%) usually occurs in the middle portion of cloudy layers, and small underestimates occur in the low and top portions of cloudy layers. The relative magnitudes of overestimates or underestimates are smaller when stratus clouds dominate in the first three and half days after the initial model spinup. [Figure 2 in Xu and Randall (1996b) shows the simulated C_s.] The differences can be as high as 50% when stratocumuli dominate at the last 30 h of the simulation. These results suggest that skewness of the distribution plays a significant role in the prediction of C_s in the stratocumulus regime. In the stratus regime, impact of skewness on the errors of predicted C_s is diminished due to the saturated mean states, that is, large positive Ω_l (see section 4).

The differences among the three PDF models are noticeable only at selected periods. The skewed and

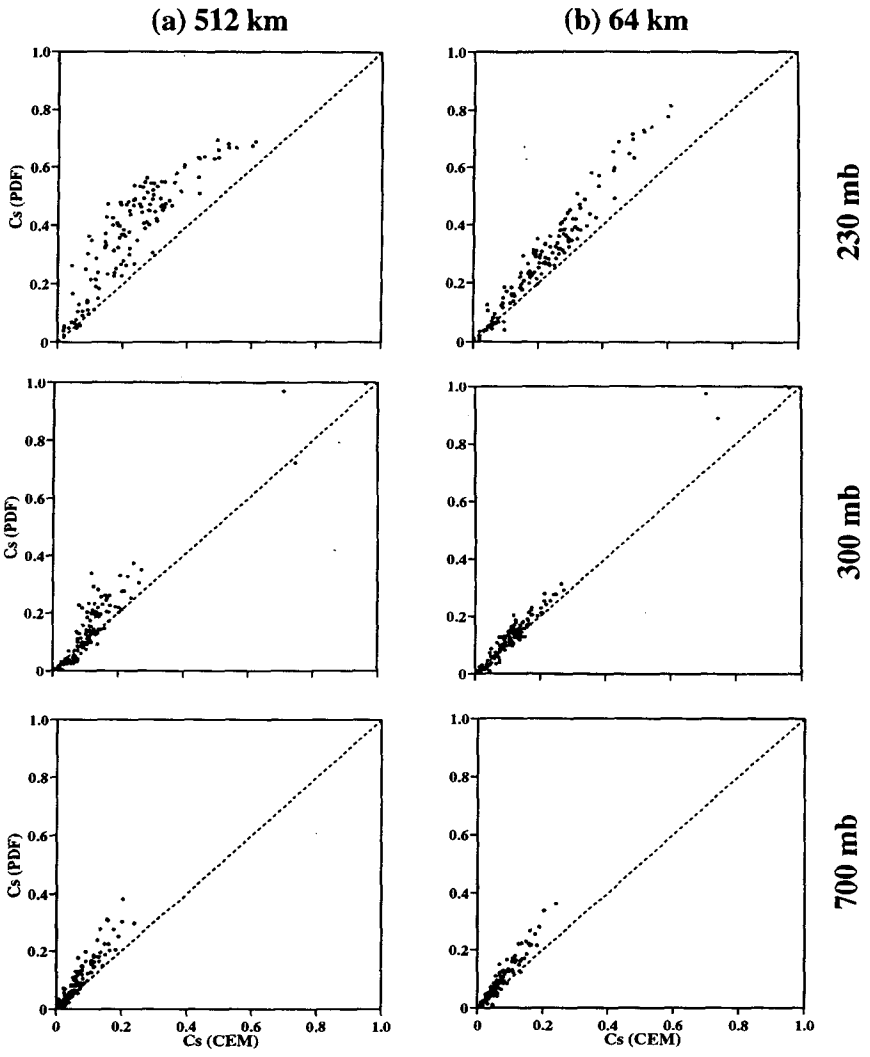

FIG. 7. Same as Fig. 5 except for the Gaussian predictions with input averaged over (a) 512-km and (b) 64-km subdomains. In (b) cloud amounts from eight subdomains are averaged.

exponential models have relatively large underestimates or overestimates when stratus clouds dominate. In stratocumulus-dominated periods, these models perform reasonably well. In summary, the Gaussian model appears superior to the other two models when stratus clouds dominate. This result is expected because of the extremely saturated mean states. The exponential and skewed models perform adequately well when stratocumuli dominate at the last 30 h of the simulation because of smaller overestimates.

Figure 9 shows similar results as in Fig. 8 except for $\bar{q}_l$. The major differences between C_s and $\bar{q}_l$ are twofold: 1) the exponential model gives smaller underestimates and overestimates, and 2) the vertical distributions of underestimate/overestimate are different. Significant overpredictions are absent throughout the simulation except for stratocumulus-dominated periods. Large underestimates appear only near the top portions of cloudy layers in the Gaussian model, while they extend to the middle and low portions of cloudy layers in the skewed model. This result again suggests that the skewed model does not perform well in the stratus regime. The reason for the good performance of the exponential model is not known. It is, however, consistent with the results obtained with the GATE data (Fig. 4).

The impact of turbulence-scale motions on predicted C_s and $\bar{q}_l$ is examined by excluding the parameterized variances of s in the definition of Ω_1. The predicted C_s usually differs by less than 1% without the turbulence-scale contributions except for the initial cloud formation (not shown). Such significant errors occur only near the cloud-base levels for a brief period during the simulation of 19 and 21 June. Therefore, conclusions

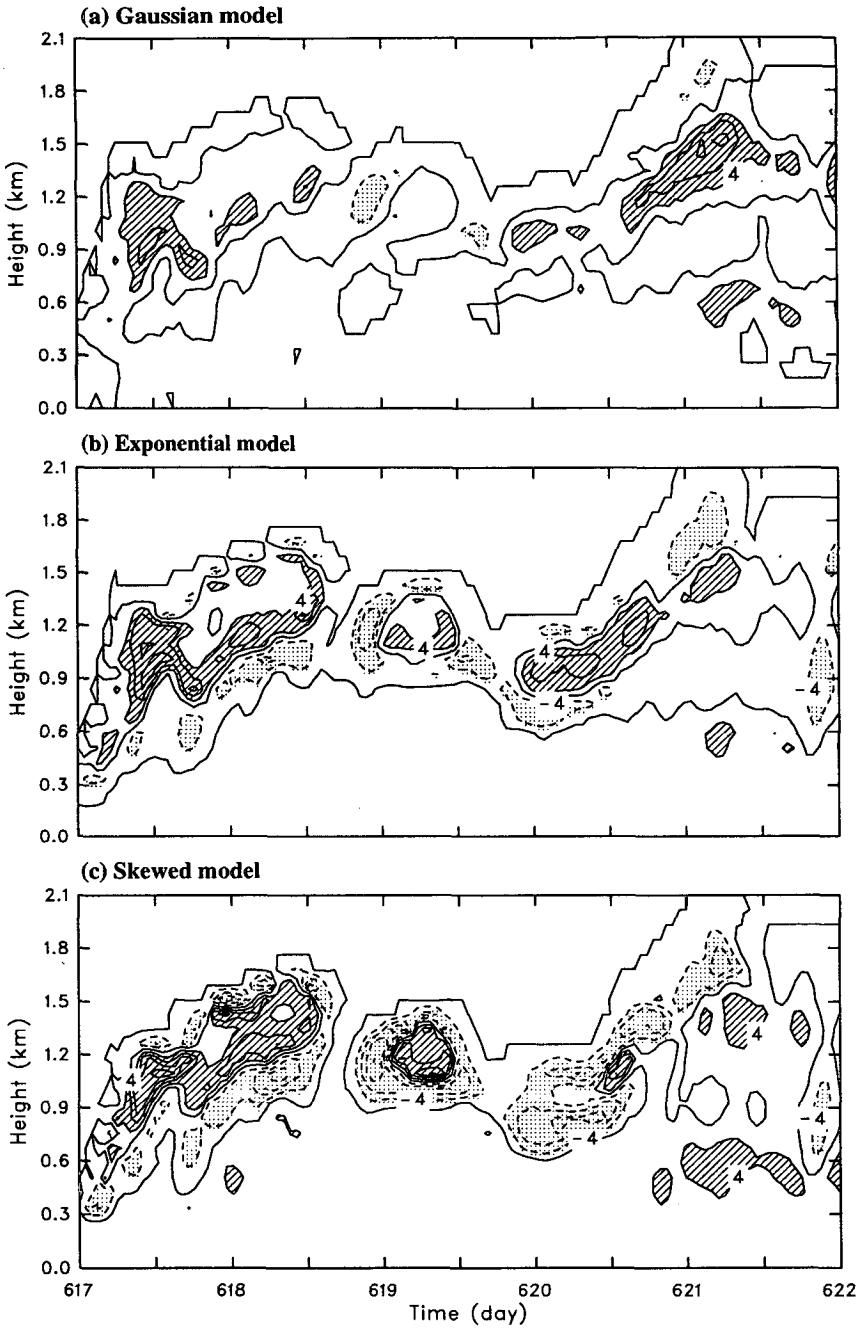

FIG. 8. Same as Fig. 2 except for the difference between predicted and simulated cloud amounts for the ASTEX simulation. The contour interval is 4%. Contours over 4% are hatched, while contours less than −4% are dotted.

on the performance of the three PDF models are not expected to be altered much.

4. Estimated PDFs from the simulated data

a. PDFs of conservative variables

Figure 10a shows the estimated PDF from the simulated GATE data. Simulated data over 18 days were used in the computation of the PDF. On each plot, the skewness factor S_s is also shown, which gives some measure of the deviation from the Gaussian PDF. The most distinct features in Fig. 10a are that 1) deviations from the Gaussian model are present at all levels and 2) S_s values are all positive. The overprediction in the low and upper troposphere, as shown in Fig. 3b, may be well interpreted by

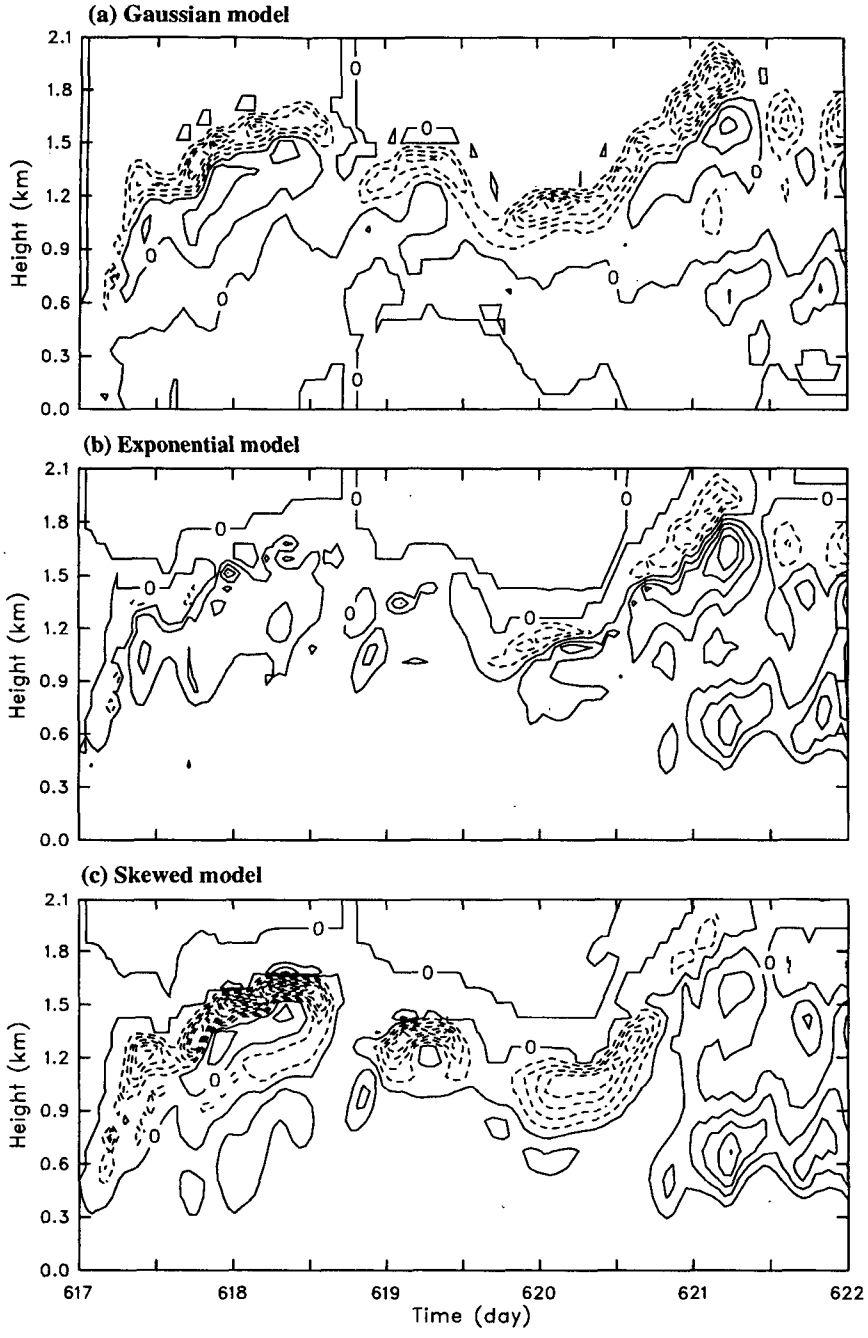

FIG. 9. Same as Fig. 2 except for the difference between predicted and simulated cloud water mixing ratio for the ASTEX simulation. The contour interval is 0.01 g kg^{-1}.

comparing the computed PDF with the Gaussian PDF [cf. (2)]. This is because C_s corresponds to the area from $-\Omega_1$ to $+\infty$. For instance, for a negative Ω_1 at 300 mb, the area under the dashed curve (Gaussian) is greater than that under the solid curve, indicating an overprediction. For a very negative Ω_1 at 470 mb (s is greater than 1.5; very dry mean state), the area under the dashed curve is less than that under the solid curve, indicating an underprediction. For a very large positive Ω_1, the overprediction in the lower and upper troposphere may be diminished due to cancellation of the areas above and under the dashed curve in Fig. 10a. The argument presented above is, however, only held for the mean C_s over 18 days (Fig. 3b).

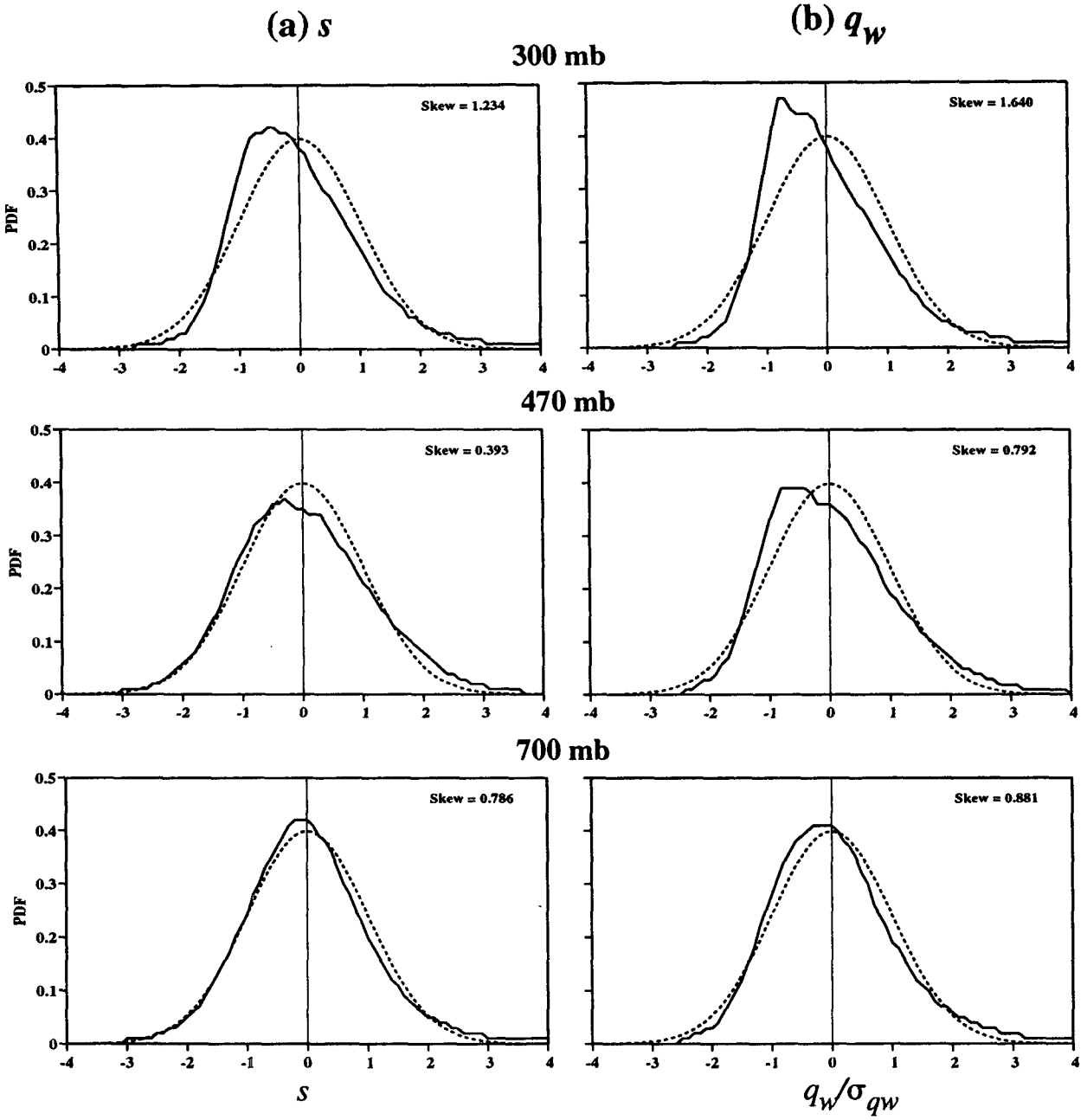

FIG. 10. Probability distribution functions of (a) s and (b) q_w (solid) at three selected levels for the 512-km domain of a GATE simulation, where s is a combination of q_w and $_l$. The dashed line is the Gaussian PDF. The skewness factor is also shown in each plot.

Note that S_s shown in Fig. 10a measures the adequacy of the Gaussian PDF. It should be pointed out that the estimated PDF is *not* based on data of one ensemble of clouds; a PDF for an individual ensemble (3-h time average) should be used to explain each scatter point shown in Figs. 5 and 6. The individual PDFs are expected to be much more skewed, which would explain the large deviations of predicted C_s and $\overline{q_l}$ from the Gaussian model shown in the left panels of Figs. 5 and 6.

Figure 11 shows scatterplots of the overprediction of C_s by the Gaussian model versus S_s at three selected levels for the 64-km subdomain. It is apparent that most skewness factors are between -0.5 and $+3$. These skewness factors are computed without including the contributions from parameterized turbulence-scale motions (less than the CEM grid size of 2 km). Since these motions are generally assumed to have PDFs close to Gaussian, their influence on the computed S_s is minor

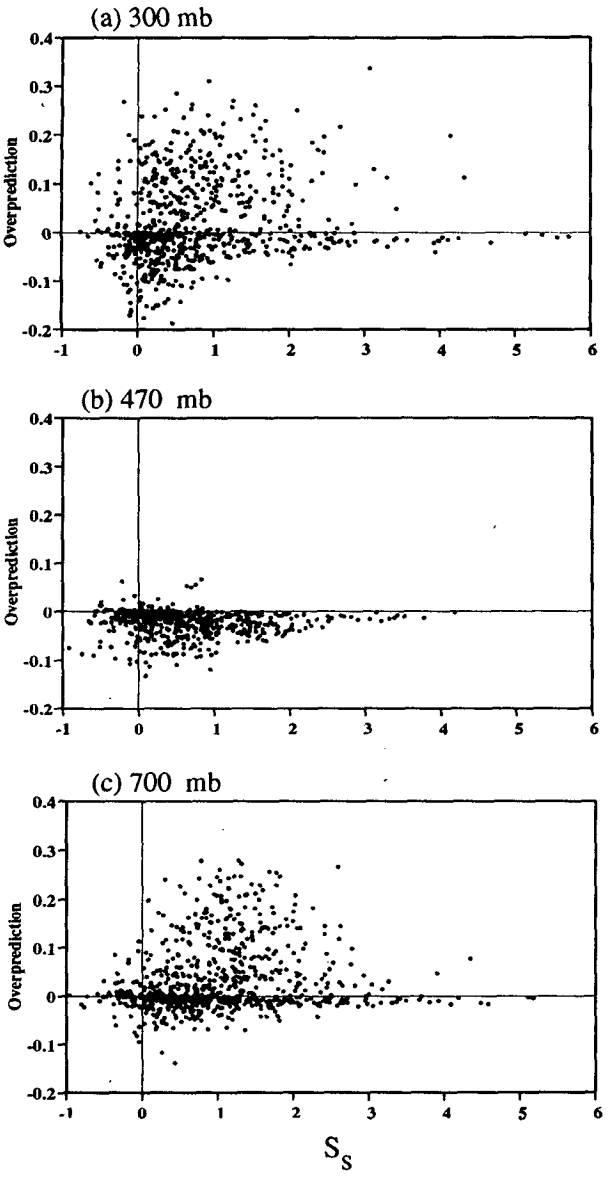

FIG. 11. The overprediction of stratiform cloud amount by the Gaussian model as a function of skewness factor S_s for the 64-km subdomain of a GATE simulation.

except for slightly altering the magnitude due to their contributions to the variance.

Another feature in Fig. 11 is that the overprediction of C_s generally increases as S_s increases (for a given mean state). The large range of overprediction for a given S_s is, however, related to different mean states of cloud ensembles (Bougeault 1982). For example, the cloud amounts corresponding to the points with extremely large S_s are very small (not shown) and the overpredictions are almost zero or negative. This is because these mean states are very dry and can only be associated with small C_s. On the other hand, underpredicted C_s shown in Fig. 11 is mainly associated with a relatively dry mean state (see the middle panel of Fig. 10a with Ω_1 greater than 1).

In summary, the three PDF models show some limited successes in the sense that their performance is highly dependent upon cloud regimes and grid sizes of GCMs. Some modifications are, however, clearly needed to incorporate skewness since a wide range of scales of motion coexist within a GCM grid cell. As mentioned earlier, the Gaussian PDF was originally intended to represent the subgrid-scale turbulence in a high resolution model. The approach adopted by Ricard and Royer (1993) for the GCM subgrid cloudiness parameterization is thus invalid because of the different nature of unresolvable-scale motions. (The dependence of PDFs on the scale of motion is addressed in the following section.) Smith (1990) used a symmetric triangular PDF, which is an oversimplification based on the present results.

Another type of PDF is based solely on q_w (Le Treut and Li 1991; Table 1). Figure 10b shows that the q_w PDF is very similar to the joint PDF of θ_l and q_w shown in Fig. 10a except for slightly larger positive S_s at all levels; the q_w PDFs are much more positively skewed in the middle and upper troposphere. The similarity of PDFs between q_w and s is related to the extremely high correlation between q'_w and s (>0.90). It may confirm the assumption used by Le Treut and Li (1991) that the θ_l variance is smaller than the q_w variance. Thus, it is expected that S_s associated with individual q_w PDFs are comparable to those of s shown in Fig. 11. In summary, results shown in Fig. 10b suggest that (i) the uniform distribution of q_w, as adopted by Le Treut and Li (1991), is not a good assumption and (ii) the q_w PDF can be used to approximate the s PDF.

b. Scale dependence of estimated PDFs

A low-pass filter (Ormsby 1961) is used to obtain skewness at certain scales of motion. The response function in this filter is 1 if the scale is greater than the roll-off termination length. Then, the response function decreases (nearly) linearly from 1 to 0 as the scale decreases from the roll-off length to the cut-off length. The response function is zero if the scale of motion is less than the cut-off length.

Figure 12 shows scatter diagrams of skewness of all scales versus skewness of limited scales at three selected levels. The roll-off termination lengths are 20 and 64 km in Figs. 12a and 12c, respectively. The cutoff lengths are half of the roll-off termination lengths. The limited scales in Fig. 12a are those eliminated in Fig. 12b.

The results in Fig. 12 show that the small scales of motion are associated with negative S_s (as mentioned earlier, inclusion of turbulence-scale motion would reduce the magnitude of S_s) while the larger scales of motion are associated with positive S_s. The larger scales of motion also tend to dominate S_s of all scales of motion. This explains why S_s of all scales of motion are rarely negative

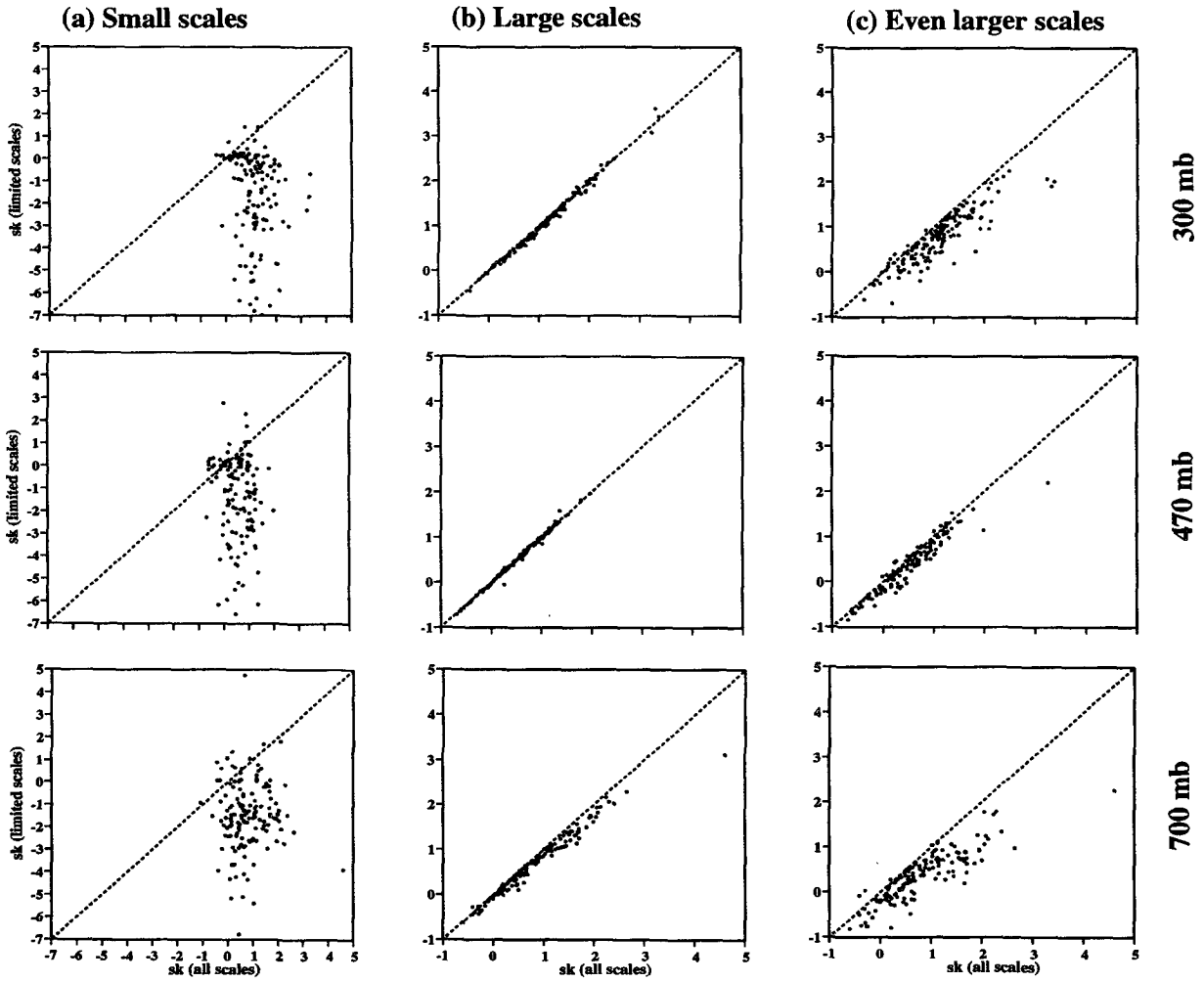

FIG. 12. Scatterplot of skewness factor of all scales vs that of limited scales. See text for explanation.

for the GATE simulation. Thus, the estimated PDFs are highly dependent upon which types of motion coexist within a grid cell of climate models. For example, negative S_s is present in the stratus regime while positive S_s appears in the cloudy layer of the stratocumulus regime during ASTEX (not shown). The scale dependency of skewness should be regarded as the most important result in explaining the PDFs in the context of climate models.

c. *PDFs of nonconservative variables*

Nonconservative variables, including the liquid-water mixing ratio $\bar{q}_l$ and vertical velocity w, have previously been used in cloudiness parameterizations (Sasamori 1975; Manton and Cotton 1977; Banta and Cotton 1980; Mocko and Cotton 1995).

The PDFs for q_l (Fig. 13a) are computed using data only with positive q_l. That is, all data with zero q_l are not used. Figure 13a shows that the PDF of q_l is exponential at all levels, that is, a highly skewed distribution. The exponential functions are slightly different in terms of their slopes among the different levels. The present results contradict the Gaussian assumption used by Manton and Cotton (1977). Nevertheless, the usefulness of such a PDF is debatable because $\bar{q}_l$ is not a conservative variable.

Figure 13b shows the PDFs for w. The distributions are very similar to the Gaussian PDF except for much higher values around $w = 0$. This feature may be related to the inclusion of cloud-free grids that may be a part of the environment of cumulus clouds. If these grids are excluded, the distributions are expected to be very close to Gaussian. This result agrees with the assumption used by Sasamori (1975).

5. Further evaluation of RH-based parameterizations

Xu and Randall (1996b) show that existing RH-based cloudiness parameterizations are not universally valid at

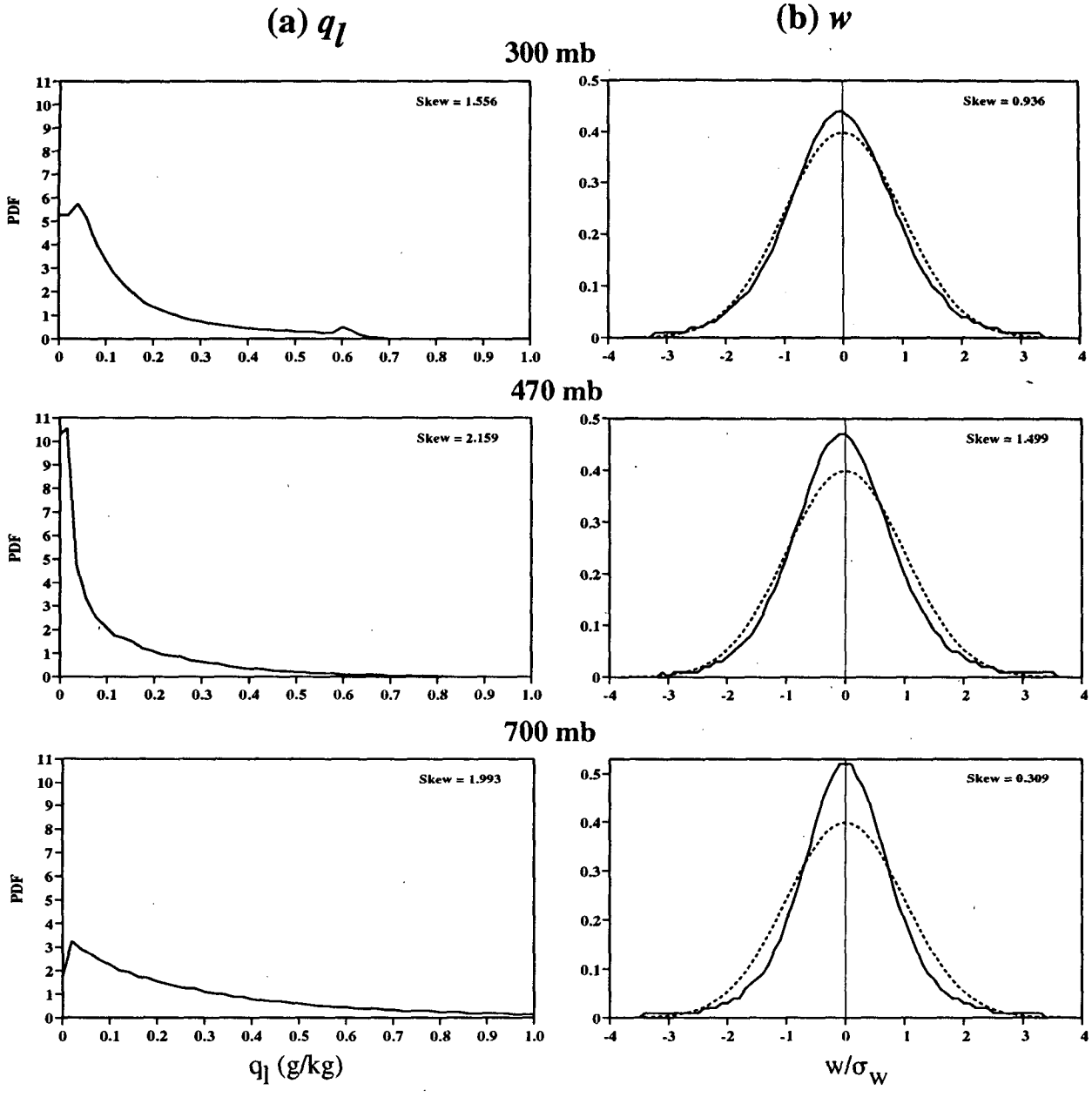

FIG. 13. Same as Fig. 10 except for the PDFs of (a) cloud water + ice mixing ratio q_l and (b) vertical velocity w/σ_w. Here q_l is not normalized.

the resolution of GCMs if RH is used as the only predictor. For example, when convective anvils are dissipated by ambient air entrainment, the cloud air mass decreases while the ambient relative humidity increases. Thus, the domain-averaged RH also increases. This indicates that the cloudiness should decrease as the RH increases. In the new parameterization proposed by Xu and Randall (1996b), the cloud amount increases as the grid-cell RH increases for a *given amount of condensate*. For the same example, the predicted cloud amount would decrease because contribution to cloudiness due to the decrease in the amount of condensate far outweighs that due to the increase of RH.

Nevertheless, a seldom examined aspect of RH-based parameterizations is whether or not there is a threshold RH for zero cloudiness. Figure 14 shows the PDFs for clear-region averaged RH over 512-km and 64-km domains. The averaged RH is termed "threshold" RH, as in Sundqvist (1978). The PDF is calculated from instantaneous threshold RHs of a given sub-

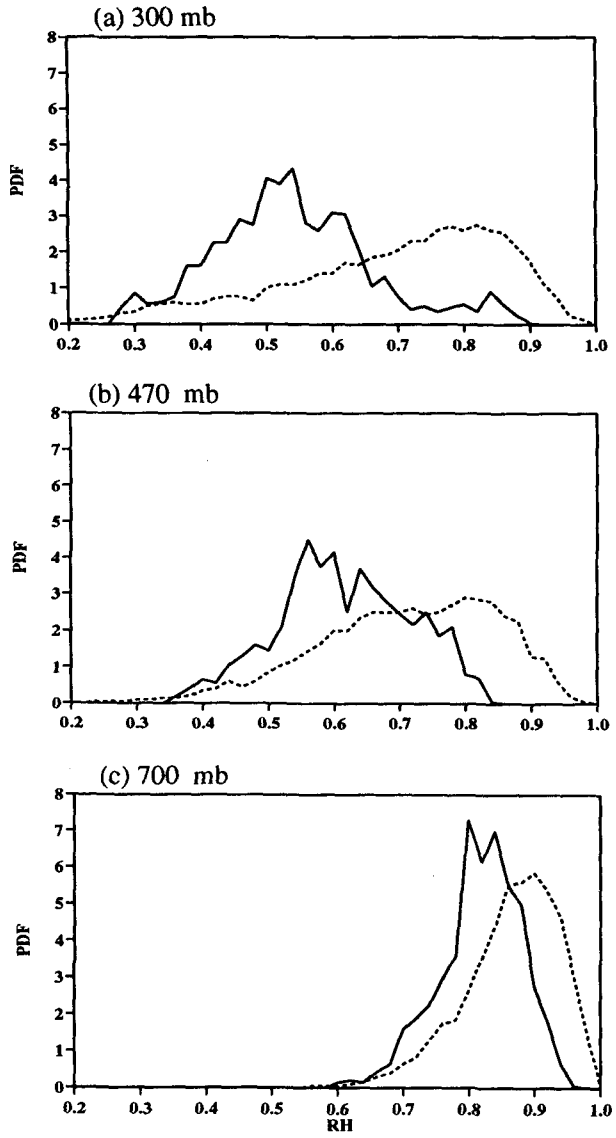

FIG. 14. Probability distribution function of clear-region averaged RHs for the 512-km (solid) and 64-km (dashed) domains at selected levels of a GATE simulation.

domain with nonzero cloudiness. Figure 14 shows that there is no *unique* threshold RH because there is no dominant and concentrated peak in the PDF. In the low troposphere, the threshold RHs are located at a narrow range (~20%). In the upper troposphere, the range of threshold RHs is very wide: between 20% and 90% for the 512-km domain. Another important feature in Fig. 14 is that the threshold RHs become larger as the horizontal-averaging distance becomes smaller, especially in the middle and upper troposphere. This is consistent with simple physical intuition.

Sundqvist et al. (1989) introduced a "basic" threshold RH so that the clear-region averaged RH is a function of cloudiness. The PDFs for such a basic threshold RH yield results similar to those shown in Fig. 14 except that the distribution functions are slightly shifted toward smaller RHs (not shown). Thus, the threshold RH introduced by Sundqvist et al. (1989) and also used by Del Genio et al. (1996) is not supported by the simulated data.

Results in Fig. 14 indicate that the assumption of a uniform RH in clear regions used in Sundqvist (1978), Sundqvist et al. (1989), Slingo (1980, 1987), and Del Genio et al. (1996) and many others is clearly too simple to represent the atmosphere. The threshold RHs may depend on the convective circulations in the stratiform regions and the intensity of cumulus convection within the subdomain. This issue needs to be further investigated. Another important result from this study is that the threshold RH is a function of the grid size of climate models, at least in the range of 64 km to 512 km. Thus, use of RH alone is highly questionable for predicting cloud amount using a unique threshold RH. This provides a motivation for the approach proposed in the companion paper (Xu and Randall 1996b). Xu and Randall (1996b) develop a "semi-empirical" parameterization using the grid-average condensate mixing ratio as the primary predictor, while RH is only used to determine the slope of the variation of C_s with respect to the grid-average condensate mixing ratio.

6. Discussion and conclusions

In this study, existing cloudiness parameterizations based on specified probability distribution functions (PDFs) and large-scale RH have been evaluated with data produced from explicit simulations of observed tropical cloud systems during GATE and subtropical stratus and stratocumulus during ASTEX.

PDF-based parameterizations were originally intended for use in cloud-resolving models where subgrid cloudiness is only associated with turbulence-scale motion (Sommeria and Deardorff 1977; Mellor 1977). In climate models, a wide range of scales of motion, from turbulence through cloud scale to mesoscale, coexist within a grid cell. In this study, the Gaussian, exponential, and positively skewed PDF models have been evaluated with numerically simulated datasets. The main conclusion is that the performance of these PDF models is highly dependent upon the cloud regimes. The Gaussian model performs reasonably well for the subtropical stratus regime, while the exponential and skewed models are slightly more adequate for predicting the subtropical stratocumulus and stratiform clouds associated with cumulus convection. Although the skewed model can better predict the cloud fraction, the liquid water mixing ratio is better predicted with the Gaussian model for the tropical stratiform regime. The performance of these models is also dependent upon the averaging distance for the input data, suggesting that existence of mesoscale motions affects the accuracy of the prediction.

Modifications to some PDF-based formulations are suggested, especially with regard to the inclusion of skewness of conservative variables. The skewness factors are found to be highly dependent upon which scales of motion coexist within a grid cell. Such modifications require that the third-moment terms must be known from a "turbulence scale" parameterization in climate models. Such macroscopic convective turbulence parameterizations are highly questionable because of the coexistence of a wide range of scales of motion within a grid cell. Therefore, PDF-based parameterizations of fractional cloudiness are not practical at present unless the high-order moments can be physically determined in climate models.

Simulated data have also been used to evaluate the existence of a unique threshold RH for zero cloudiness, as widely used in existing parameterizations (e.g., Smagorinsky 1960; Sundqvist 1978; Sundqvist et al. 1989; Smith 1990; Walcek 1994; Del Genio et al. 1996). It is found that the clear-region averaged RHs can vary greatly with height and the grid size of climate models, in addition to their wide range of variations at a given height. Thus, RH-based parameterizations are not readily supported by results presented in this study.

This study points out the need for other types of approaches in the cloudiness parameterization problem. Within the framework of prognostic parameterization of cloud formation and dissipation processes as pioneered by Sundqvist (1978), inclusion of grid-average condensate mixing ratio as a predictor is a logical step for diagnosing fractional cloudiness. Xu and Randall (1996b) have proposed such a framework in the companion paper. Another approach is to prognostically determine the cloudiness as proposed by Tiedtke (1993), although those plausible formulations of sources and sinks for cloud water and cloud fraction need to be verified with observations or evaluated against explicit simulation of cloud ensembles.

Acknowledgments. This research was primarily supported by the Environmental Sciences Division of the U.S. Department of Energy under Grant DE-FG03-95ER61968 as part of the Atmospheric Radiation Measurement Program. It was also partially supported by CERES under Grant NAG1-1266 and FIRE under Grant NAG1-1701, both from NASA to the Colorado State University.

Observational datasets used in the present study were kindly provided by Professor R. Reed of the University of Washington, Professor M. Yanai of UCLA for the GATE sounding data, and Professor C. Bretherton of the University of Washington for the ASTEX data. We are grateful to Professor A. Arakawa and two anonymous reviewers for their constructive comments, Tammy Weckwerth for improving the manuscript, and Bruce Wielicki and Steve Krueger for their interest in this work.

REFERENCES

Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *J. Atmos. Sci.*, **31,** 674–701.

Banta, R., and W. R. Cotton, 1980: On computing average cloud-water quantities in a partially cloudy region. *J. Rech. Atmos.*, **14,** 249–259.

Bougeault, P., 1981: Modeling the trade-wind cumulus boundary layer. Part I: Testing the ensemble cloud relations against numerical data. *J. Atmos. Sci.*, **38,** 2414–2428.

——, 1982: Cloud-ensemble relations based on the Gamma probability distribution for the higher-order models of planetary boundary layer. *J. Atmos. Sci.*, **39,** 2691–2700.

Del Genio, A. D., M.-S. Yao, W. Kovari, and K. K.-W. Lo, 1996: A prognostic cloud water parameterization for global climate models. *J. Climate,* **9,** 270–304.

Gregory, D., and M. J. Miller, 1989: A numerical study of the parameterization of deep tropical convection. *Quart. J. Roy. Meteor. Soc.*, **115,** 1209–1241.

Krueger, S. K., 1988: Numerical simulation of tropical cumulus clouds and their interaction with the subcloud layer. *J. Atmos. Sci.*, **45,** 2221–2250.

Le Treut, H., and Z.-X. Li, 1988: Using meteosat data to validate a prognostic cloud generation scheme. *Atmos. Res.*, **21,** 273–292.

——, and ——, 1991: Sensitivity of an atmospheric general circulation model to prescribed SST changes: Feedback effects associated with the simulation of cloud optical properties. *Climate Dyn.*, **5,** 175–187.

Lipps, F. B., and R. S. Hemler, 1986: Numerical simulation of deep tropical convection associated with large-scale convergence. *J. Atmos. Sci.*, **43,** 1796–1816.

Manton, M. J., and W. R. Cotton, 1977: Formulation of approximate equations for modeling moist deep convection on the mesoscale. Atmospheric Science Paper 266, 42 pp. [Available from Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523.]

Mellor, G. L., 1977: The Gaussian cloud model relations. *J. Atmos. Sci.*, **34,** 356–358.

Mocko, D. M., and W. R. Cotton, 1995: Evaluation of fractional cloudiness parameterizations for use in a mesoscale model. *J. Atmos. Sci.*, **52,** 2884–2901.

Ormsby, J. F. A., 1961: Design of numerical filters with applications to missile data processing. *J. Assoc. Comput. Mach.*, **8,** 440–466.

Ricard, J. L., and J. F. Royer, 1993: A statistical cloud scheme for use in an AGCM. *Ann. Geophys.*, **11,** 1095–1115.

Sasamori, T., 1975: A statistical model for stationary atmospheric cloudiness, liquid water content, and rate of precipitation. *Mon. Wea. Rev.*, **103,** 1037–1049.

Slingo, J. M., 1980: A cloud parameterization scheme derived from GATE data for use with a numerical model. *Quart. J. Roy. Meteor. Soc.*, **106,** 747–770.

——, 1987: The development and verification of a cloud prediction scheme for the ECMWF model. *Quart. J. Roy. Meteor. Soc.*, **113,** 899–927.

Smagorinsky, J., 1960: On the dynamical prediction of large-scale condensation by numerical methods. *Physics of Precipitation, Geophys. Monogr.*, No. 5, Amer. Geophys. Union, 71–78.

Smith, R. N. B., 1990: A scheme for predicting layer clouds and their water content in a GCM. *Quart. J. Roy. Meteor. Soc.*, **116,** 435–460.

Sommeria, G., and J. W. Deardorff, 1977: Subgrid-scale condensation in models of nonprecipitating clouds. *J. Atmos. Sci.*, **34,** 344–355.

Soong, S.-T., and W.-K. Tao, 1980: Response of deep tropical cumulus clouds to mesoscale processes. *J. Atmos. Sci.*, **37,** 2016–2034.

Sundqvist, H., 1978: A parameterization scheme for non-convective condensation, including prediction of cloud water content. *Quart. J. Roy. Meteor. Soc.*, **104,** 677–690.

——, E. Berge, and J. E. Krishjansson, 1989: Condensation and cloud parameterization studies with a mesoscale numeri-

cal weather prediction model. *Mon. Wea. Rev.*, **117,** 1641–1657.

Tiedtke, M., 1993: Representation of clouds in large-scale models. *Mon. Wea. Rev.*, **121,** 3040–3061.

Walcek, C. J., 1994: Cloud cover and its relationship to relative humidity during a springtime midlatitude cyclone. *Mon. Wea. Rev.*, **122,** 1021–1035.

Xu, K.-M., 1995: Partitioning mass, heat, and moisture budgets of explicitly simulated cumulus ensembles into convective and stratiform components. *J. Atmos. Sci.*, **52,** 551–573.

——, and S. K. Krueger, 1991: Evaluation of cloudiness parameterizations using a cumulus ensemble model. *Mon. Wea. Rev.*, **119,** 342–367.

——, and D. A. Randall, 1996a: Explicit simulation of cumulus ensembles with the GATE Phase III data: Comparison with observations. *J. Atmos. Sci.*, in press.

——, and ——, 1996b: A semiempirical cloudiness parameterization for use in climate models. *J. Atmos. Sci.*, **53,** 3084–3102.

Yamada, T., and G. L. Mellor, 1979: A numerical simulation of BOMEX data using a turbulence closure model coupled with ensemble cloud relations. *Quart. J. Roy. Meteor. Soc.*, **105,** 915–944.

Yanai, M., S. Esbensen, and J.-H. Chu, 1973: Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *J. Atmos. Sci.*, **30,** 611–627.